Adaptive Beamforming for Microphone Arrays on Seat Belts

Mohamed Krini[1], Vasudev Kandade Rajan[2], Klaus Rodemer[1], Gerhard Schmidt[2]

[1] *paragon AG, Division Acoustics, Schwalbenweg 29, Delbrück, Germany*

E-Mail: mohamed.krini/klaus.rodemer@paragon.ag

[2] *Christian-Albrechts-Universität zu Kiel, Digital Signal Processing and System Theory, Kaiserstr. 2, Kiel, Germany*

E-Mail: vakr/gus@tf.uni-kiel.de

Abstract

Belt-microphones are an interesting alternative to conventional microphones used for hands-free telephony or speech dialogue systems in automobile environments. An enhanced signal quality in terms of high signal-to-noise ratio (SNR) can usually be reached compared to other microphone positions commonly placed, e.g., at the rear view mirror, the steering wheel, or the center console. The seat belt-microphone system consists of three microphones which are placed around the shoulder and chest of a sitting passenger. The entire geometry is flexible and is likely to change due to movements of the passenger. From the arrangement of these three microphones, that microphone is often selected which has the best overall signal quality in terms of high SNR. Further improvements can be achieved if all microphone signals are combined to a single output signal. In this contribution the signal combination is performed with an adaptive beamformer. The overall performance of the proposed beamformer designed for belt-microphones is analyzed in terms of SNR and SIR (signal-to-interference ratio) and compared with that of the single best microphone.

Seat Belt-Microphones

The seat belt-microphones are manufactured by the company paragon AG [1]. In Fig. 1 an example of a belt-microphone system installed in a car is shown. It consists of three omnidirectional microphones spaced by 160 mm. The backside of the belt has an even surface and all signal lines needed for signal transmission and voltage supply are integrated into the seat belt. A sophisticated mounting technique for the microphone capsules has been developed. The signal lines are made of a special alloy and a flexible structure of wires are weaved into the seat belts so that they appear invisible. It has been proven that the safety, usability, and comfort of a seat belt with integrated microphones are still maintained. The microphones satisfy the VDA 1.5 specifications and receive signals between 100 Hz–8 kHz.

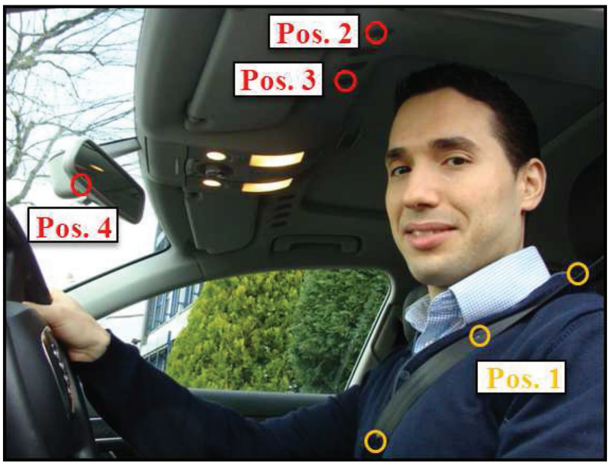

Figure 1: Belt-microphones (Pos. 1) and microphones positioned at the roof and at the mirror (Pos. 2 – 4).

In Fig. 2 the belt-microphone system is compared with three hands-free microphones placed at different positions (see Fig. 1) in terms of the average SNR for driving speeds between 120 and 160 km/h. The distances from different microphone positions to the mouth are: $20 - 27$ cm (Pos. 1), 28 cm (Pos. 2/3), 58 cm (Pos. 4). All microphones are calibrated to have the same power at stand-still. This comparison shows that at higher frequencies the behavior of all microphones is more or less similar. Whereas at low and medium frequencies, the belt-microphone outperforms conventional hands-free microphones. An improvement of up to $6 - 10$ dB in SNR can be achieved.

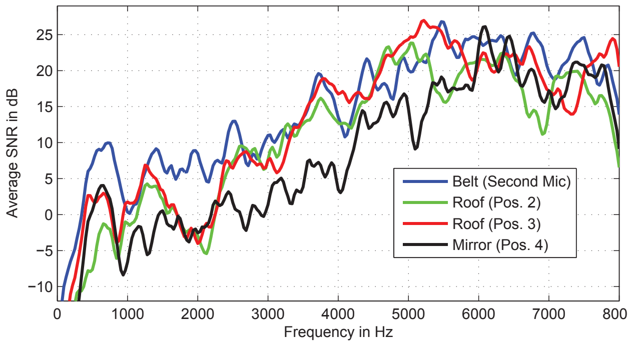

Figure 2: SNR measured at different microphone positions.

Processing Framework

The processing of the microphone signals is performed in the subband domain. The discrete-time samples of the microphone signals, which are sampled at frequency of $f_s = 16\,\text{kHz}$, are transformed into subbands by a short-term Fourier transform (STFT) [5] also referred as the analysis filterbank. The idea is to obtain a real-time processing framework for the processing of speech signals. A frame shift of $r = 128$ samples and an FFT order of $N_{\text{FFT}} = 512$ are chosen and the samples are weighted with a Hann window. The output of the analysis filterbank contains the spectra of the $M = 3$ microphone signals $Y_l(\mu, k)$ where the index $l \in \{0, 1, 2\}$ is the micro-

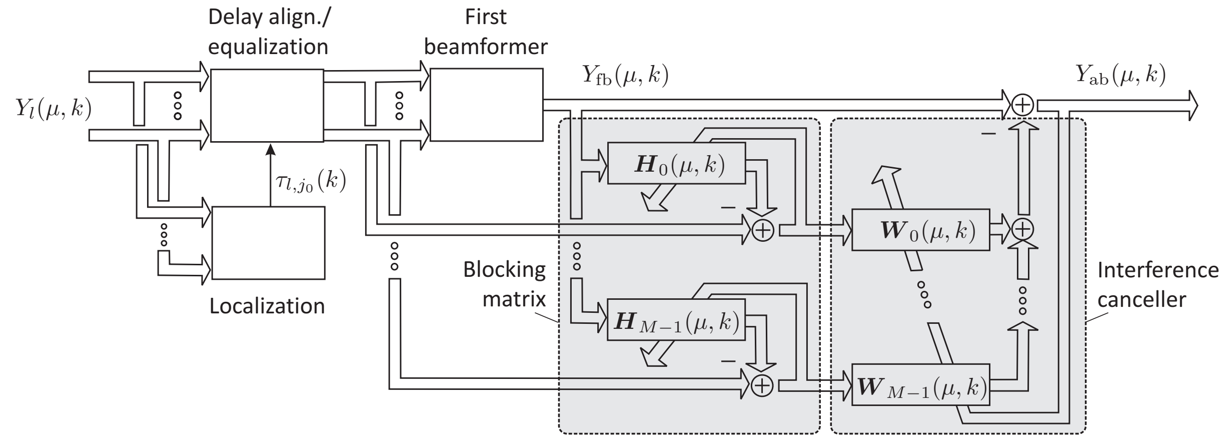

Figure 3: Overview of the proposed adaptive beamformer for seat belt-microphones.

phone index, $\mu \in \{0, 1, \ldots, N_{\text{FFT}} - 1\}$[1] is the subband index, and k is the frame index.

Adaptive Beamforming for Seat Belts

In the following we will describe details about all components that are necessary to perform a robust beamforming approach with seat belt-microphones.

Localization

Before proceeding with combining the three microphone spectra, they are pre-processed with two blocks namely the *Delay Alignment and Equalization* and the *Localization* as shown in Fig. 3. The delay alignment is performed in order to compensate for the differences in terms of distance (seen as a delay) between the mouth of the speaker and the individual microphones. The delay is estimated by a Generalized Cross-Correlation (GCC) function presented in [6]. Here a pairwise instantaneous Cross Power Spectral Density (CPSD) $\tilde{S}_{y_l,y_j}(\mu,k) = Y_l(\mu,k)\,Y_j^*(\mu,k)$ between the microphone spectra is determined for $l \neq j$. The CPSD is smoothed to avoid the jumps and variations seen on the instantaneous spectra through a first order IIR filter given by

$$\overline{S}_{y_l,y_j}(\mu,k) = \alpha_Y\,\tilde{S}_{y_l,y_j}(\mu,k) + (1-\alpha_Y)\,\overline{S}_{y_l,y_j}(\mu,k-1), \tag{1}$$

where the smoothing constant α_Y is chosen to be around 0.8. The delay is computed by the argument that maximizes the inverse Fourier transform of that quantity. Before transforming it a weighting can be applied. We utilized here the so called Phase Transformation (PHAT) [7] leading to the following inverse cross correlation sequence:

$$s_{y_l,y_j}(\kappa,k) = \frac{1}{N_{\text{FFT}}} \sum_{\mu=0}^{N_{\text{FFT}}-1} \frac{\overline{S}_{y_l,y_j}(\mu,k)}{\left|\overline{S}_{y_l,y_j}(\mu,k)\right|}\, e^{j\frac{2\pi\mu}{N_{\text{FFT}}}\kappa}. \tag{2}$$

The time-domain transformed samples are indexed by κ. Optionally a lower FFT size can be used to reduce the computational complexity by dropping bins above, e.g., 3500 Hz. The robustness can be improved by setting the lowest bins to zero before applying the transform. The maximum distance between the seat belt-microphones is 320 mm and hence the delay is limited to this distance given by $\tau_{\max}$. From the time-domain transformed frame the delay $\tau_{l,j}(\kappa)$ is computed by finding the argument κ that maximizes the cross correlation function $s_{y_l,y_j}(\kappa,k)$:

$$\tau_{l,j}(\kappa) = \underset{-\tau_{\max}<\kappa<\tau_{\max}}{\operatorname{argmax}} \left\{ s_{y_l,y_j}(\kappa,k) \right\}. \tag{3}$$

Signal Equalization

The nature of the seat belt-microphones is such that they usually pickup slightly varied ambient noise even if they are in the same environment. To achieve good combination performance it is important to correct this by equalizing the noise for all the microphones. This is achieved by using a simple multiplicative constant based noise PSD estimation for each microphone given by

$$\hat{B}_l(\mu,k) = \begin{cases} \delta_{\text{inc}} \cdot \hat{B}_l(\mu,k-1), \\ \qquad\qquad \text{if } \overline{Y}_l(\mu,k) > \hat{B}_l(\mu,k-1), \\ \delta_{\text{dec}} \cdot \hat{B}_l(\mu,k-1), \text{else}, \end{cases} \tag{4}$$

where $\hat{B}_l(\mu,k)$ is the estimated magnitude spectrum of background noise for each microphone, δ_{inc} is the incremental constant, and δ_{dec} is the decremental constant, with $0 \ll \delta_{\text{dec}} < 1 < \delta_{\text{inc}}$. $\overline{Y}_l(\mu,k)$ is a smoothed version of the magnitude of the spectrum $\widetilde{Y}_l(\mu,k)$ as opposed to a complex smoothed spectra obtained similarly as shown in Eq. (1). A slowly varying equalization factor $E_l(\mu,k)$ per microphone is computed and tracked based on the average background noise $\hat{B}_{\text{avg}}(\mu,k) = 1/M \sum_{l=0}^{M-1} \hat{B}_l(\mu,k)$ of all the three microphones given by

$$E_l(\mu,k) = \begin{cases} \delta_{\text{gain-inc}} \cdot E_l(\mu,k-1), \\ \qquad\qquad \text{if } \hat{B}_{\text{avg}}(\mu,k) > \hat{B}_l(\mu,k), \\ \delta_{\text{gain-dec}} \cdot E_l(\mu,k-1), \quad \text{else}. \end{cases} \tag{5}$$

The equalization factor, which is bounded by a maximum and a minimum value for safety reasons, is applied to the microphone spectra along with the estimated delay to obtain the pre-processed spectra on which the beamforming technique is applied. This is performed by

$$\widetilde{Y}_j(\mu,k) = E_j(\mu,k) \cdot Y_j(\mu,k)\, e^{-j\frac{2\pi\mu}{N_{\text{FFT}}}\tau_{j,j_0}(k)}. \tag{6}$$

Usually the center microphone is used as a (delay) reference, meaning that we use $j_0 = 1$ in Eq. 6.

[1] Since the input signals are assumed to be real, it is sufficient to store only the first $N_{\text{FFT}}/2+1$ frequency supporting points.

Revisiting SNR-based Weighting Beamformer

In the literature, e.g. in [2], SNR-based beamforming is widely presented referred to here as the *first beamformer* (see Fig. 3). In this section the SNR weighted beamformer is briefly presented so as to act as a precursor to the adaptive technique presented in the next section. The SNR-based weighting beamformer is a modified filter-and-sum beamformer which means that each input to the beamformer will be filtered and the sum of all the filtered inputs forms the output of the beamformer. In the context of this paper the inputs to the beamformer are the three equalized and delay-aligned seat belt-microphone spectra. This is shown in Eq. (7)

$$Y_{\mathrm{fb}}(\mu,k)=\sum_{l=0}^{M-1} G_l(\mu,k)\,\widetilde{Y}_l(\mu,k). \tag{7}$$

The filter $G_l(\mu,k)$ is a function of the normalized SNR computed per subband of the respective microphone. The SNR per subband $\Gamma_l(\mu,k)$ is computed by

$$\Gamma_l(\mu,k)=\frac{\left|\widetilde{Y}_l(\mu,k)\right|^2}{\hat{B}_l^2(\mu,k)}. \tag{8}$$

The normalized SNR is computed by dividing per subband SNR by the sum of all the subband SNRs of three microphones given by

$$\widetilde{\Gamma}_l(\mu,k)=\frac{\Gamma_l(\mu,k)}{\sum\limits_{j=0}^{M-1}\Gamma_j(\mu,k)}. \tag{9}$$

Since the short-term SNR of the individual microphone signals is highly varying, the filter function is computed by a smoothed version of the normalized SNRs. In addition, the filter function should be updated only during speech activity. The smoothing is again performed by an IIR filter with the smoothing constant switching between a constant and 0 to ensure that the previous values are applied during non-speech frames. This is captured in Eqs. (10) and (11)

$$G_l(\mu,k)=(1-\alpha_l(k))\,G_l(\mu,k-1)+\alpha_l(k)\,\widetilde{\Gamma}_l(\mu,k), \tag{10}$$

where

$$\alpha_l(k)=\begin{cases}\alpha_{\mathrm{SNR}}, & \text{if } \frac{1}{N_{\mathrm{FFT}}}\sum\limits_{\mu=0}^{N_{\mathrm{FFT}}-1}\Gamma_l(\mu,k)>T_{\mathrm{SNR}},\\ 0, & \text{else,}\end{cases} \tag{11}$$

where α_{SNR} is the smoothing constant, and T_{SNR} is the SNR threshold parameter for voice activity detection.

Adaptive Blocking Matrix

The adaptive blocking matrix (ABM) generates a noise reference for the interference canceller (IC). A fixed blocking matrix, which subtracts adjacent equalized and time-aligned microphone subband signals, is not suitable for belt-microphones due to the strong microphone SNR variations. The ABM subtracts adaptively filtered versions of the first beamformer output $Y_{\mathrm{fb}}(\mu,k)$ from each channel input $\widetilde{Y}_l(\mu,k)$ and provides the noise reference signals $U_l(\mu,k)$ for the IC with $l\in[0,M-1]$. The SNR differences between belt-microphones and the mismatch of the steering direction can be compensated. Filters of the blocking matrix are adapted using the NLMS algorithm:

$$\boldsymbol{H}_l(\mu,k+1)=\boldsymbol{H}_l(\mu,k)+\beta_{\mathrm{bm}}^{(\mu)}(k)\,\frac{\boldsymbol{Y}_{\mathrm{fb}}(\mu,k)\,U_l^*(\mu,k)}{\|\boldsymbol{Y}_{\mathrm{fb}}(\mu,k)\|^2}\ , \tag{12}$$

where $\boldsymbol{H}_l(\mu,k)=[H_l(\mu,k),...,H_l(\mu,k-N_{\mathrm{bm}}+1)]^{\mathrm{T}}$ denote the subband filter coefficients and N_{bm} is the filter length. The vector $\boldsymbol{Y}_{\mathrm{fb}}(\mu,k)=[Y_{\mathrm{fb}}(\mu,k),...,Y_{\mathrm{fb}}(\mu,k-N_{\mathrm{bm}}+1)]^{\mathrm{T}}$ comprises the current and the last $N_{\mathrm{bm}}-1$ subband outputs of the first beamformer. Filters of the blocking matrix are only adapted if speech is picked up from the steering direction. For robustness, the filter coefficients can be limited (in terms of their magnitudes) by an upper and lower threshold [8]. The step-size $\beta_{\mathrm{bm}}^{(\mu)}(k)$ is used to control the speed of the adaptation in every subband.

Interference Canceller

The subband signals $U_l(\mu,k)$ are passed to the IC which adaptively removes the signal components that are correlated to the interference input signals from the beamformer output $Y_{\mathrm{fb}}(\mu,k)$. The adaptive filters $\boldsymbol{W}_l(\mu,k)=[W_l(\mu,k),...,W_l(\mu,k-N_{\mathrm{ic}}+1)]^{\mathrm{T}}$ of the IC are not adapted if speech is coming from the steering direction to avoid signal cancellation. N_{ic} is denoting the filter length. For filter adaptation the NLMS algorithm is used:

$$\boldsymbol{W}_l(\mu,k+1)=\boldsymbol{W}_l(\mu,k)+\beta_{\mathrm{ic}}(k)\,\frac{\boldsymbol{U}_l(\mu,k)\,Y_{\mathrm{ab}}^*(\mu,k)}{\sum_{l=0}^{M-1}\|\boldsymbol{U}_l(\mu,k)\|^2}\ . \tag{13}$$

The vector $\boldsymbol{U}_l(\mu,k)=[U_l(\mu,k),...,U_l(\mu,k-N_{\mathrm{ic}}+1)]^{\mathrm{T}}$ comprises the current and the last $N_{\mathrm{ic}}-1$ output signals of the ABM. The adaptive beamformer output is determined by $Y_{\mathrm{ab}}(\mu,k)=Y_{\mathrm{fb}}(\mu,k)-\sum_{l=0}^{M-1}\boldsymbol{U}_l^{\mathrm{T}}(\mu,k)\,\boldsymbol{W}_l(\mu,k)$. In order to increase the robustness of the beamformer the norm of the adaptive filter coefficients can be limited [8]. The control of the step-size $\beta_{\mathrm{ic}}(k)$ is described in the following section.

Adaptation Control

The step-sizes for the ABM and the IC are controlled based on the speech activity from the steering direction. As a measure for the speech activity, a ratio of the smoothed short-term powers $S_{y_{\mathrm{fb}}y_{\mathrm{fb}}}^{(\mu)}(k)$ and $S_{uu}^{(\mu)}(k)$ of the first beamformer and of the ABM output, respectively, averaged over a certain frequency range, is used:

$$r_{\mathrm{SD}}(k)=\frac{\sum\limits_{\mu=N_{\mathrm{u}}}^{N_{\mathrm{o}}} S_{y_{\mathrm{fb}}y_{\mathrm{fb}}}^{(\mu)}(k)}{\sum\limits_{\mu=N_{\mathrm{u}}}^{N_{\mathrm{o}}} S_{uu}^{(\mu)}(k)}\ . \tag{14}$$

The short-term powers are smoothed through a first order IIR-filter:

$$S_{y_{\mathrm{fb}}y_{\mathrm{fb}}}^{(\mu)}(k)=(1-\alpha)\,S_{y_{\mathrm{fb}}y_{\mathrm{fb}}}^{(\mu)}(k-1)+\alpha\,|Y_{\mathrm{fb}}(\mu,k)|^2, \tag{15}$$

$$S_{uu}^{(\mu)}(k) = (1-\alpha)\, S_{uu}^{(\mu)}(k-1) + \alpha\, \beta(\mu,k) \sum_{l=0}^{M-1} |U_l(\mu,k)|^2 .$$

The smoothing constant is chosen around $\alpha = 0.5$, the lower and upper frequencies used in Eq. 14 were set by $\Omega_{N_\mathrm{u}} = 1$ kHz and $\Omega_{N_\mathrm{o}} = 6$ kHz. $\beta(\mu,k)$ is controlled such that in periods of stationary background noise the ratio $r_\mathrm{SD}(k)$ becomes one. Only high values of $r_\mathrm{SD}(k)$ indicate signal energy from the steering direction. Thus, the filters of the ABM are adjusted only when $r_\mathrm{SD}(k)$ exceeds a predetermined threshold $t_\mathrm{bm} = 0.3$ using:

$$\beta_\mathrm{bm}^{(\mu)}(n) = \begin{cases} \beta_\mathrm{bm}^{(\max)}, & \text{if } r_\mathrm{sd}(n) \geq t_\mathrm{bm} \\ & \wedge\, S_{b_\mathrm{fb} b_\mathrm{fb}}^{(\mu)}(k)\, K < |Y_\mathrm{fb}(\mu,k)|^2, \\ 0\,, & \text{else}\,, \end{cases} \quad (16)$$

where $S_{b_\mathrm{fb} b_\mathrm{fb}}^{(\mu)}(k)$ denotes the estimated PSD of the noise at the first beamformer output and K is set to 6 dB. The adaptive filters of the IC are controlled using:

$$\beta_\mathrm{ic}(n) = \begin{cases} \beta_\mathrm{ic}^{(\max)}\,, & \text{if } r_\mathrm{sd}(n) < t_\mathrm{ic}\,, \\ 0\,, & \text{else}\,, \end{cases} \quad (17)$$

with $t_\mathrm{ic} = 0.2$. The maximum step-sizes can be set to $\beta_\mathrm{ic}^{(\max)} = 0.1$ and $\beta_\mathrm{bm}^{(\max)} = 0.2$.

Experimental Results

For evaluating the performance of the beamformer real world recordings have been made at a speed of 120 km/h using the driver belt. For analysis, two belt-microphones with high SNRs have been used to reduce computational complexity. A frequency selective SNR and SIR analysis has been made to compare the non-adaptive and the adaptive beamformers with the best single microphone. The results of the SNR can be seen in Fig. 4. The SNR performance of the first beamformer is slightly better than that of the single best microphones (on average about 2 dB). The overall SNR performance can be further increased by about 5 dB when using the proposed adaptive beamformer compared to the best belt-microphone. The SIR analysis as shown in Fig. 5 indicates that on overage about 2 dB SIR-improvement with the first beamformer and 6 dB with the adaptive beamformer can be achieved. The proposed adaptive beamformer is suitable for highly suboptimal array geometries and shows robust performance when the signal source position is changing fast.

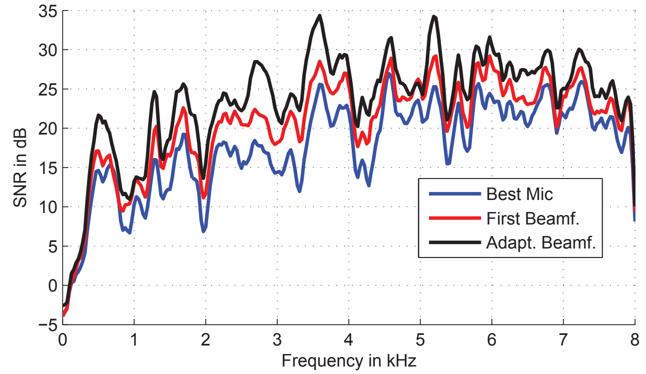

Figure 4: SNR comparison between the best belt-microphone and the beamformer outputs.

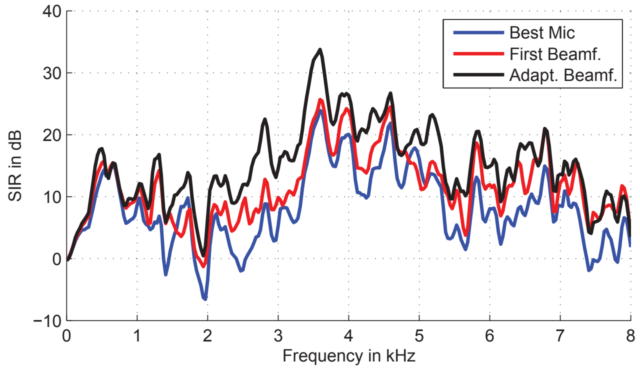

Figure 5: SIR comparison between the best belt-microphone and the beamformer outputs.

Conclusion

In this contribution an adaptive beamformer for seat belt-microphones was presented. A single unprocessed belt-microphone already achieves a better SNR compared to microphones placed at common positions in a vehicle. Since the entire array geometry may change, a reliable and robust localization method was implemented. Highly varying signal power at different belt-microphones can be compensated with an adaptive blocking matrix. A robust measure was used to effectively discriminate between desired and interference speech. Evaluations show that the SNR and SIR can be enhanced with an adaptive beamformer for belt-microphone systems although the farfield condition is not satisfied and the entire array geometry is not fixed. The performance, especially in terms of SIR, can be further increased by using a spatial filter as a post-processor to an adaptive beamformer.

References

[1] paragon AG, URL: `http://www.paragon.ag/`.

[2] V.K. Rajan, S. Rohde, G. Schmidt, J. Withopf, "Signal Processing for Microphone Arrays on Seat Belts", Workshop on DSP for In-Vehicle Systems, Seoul, Korea, 2013.

[3] Y. Ephraim and D. Malah, "Speech Enhancement using a MMSE Short-Time Spectral Amplitude Estimator, IEEE Trans. on ASSP, vol. 32, no. 6, pp. 1109-1121, 1984.

[4] M. Krini, K. Rodemer, "Seat Belt-Microphone Systems and their Application to Speech Signal Enhancement", 40th Annual German Congress on Acoustics (DAGA'14), Oldenburg, Germany, 2014.

[5] Allen, J.B., "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform", Acoustics, Speech and Signal Processing, IEEE Transactions on , vol.25, no.3, pp.235,238, Jun 1977.

[6] Knapp, C.; Carter, G.Clifford, "The Generalized Correlation Method for Estimation of Time Delay", Acoustics, Speech and Signal Processing, IEEE Transactions on , vol.24, no.4, pp.320,327, Aug 1976

[7] Carter, G.Clifford; Nuttall, Albert H.; Cable, P., "The Smoothed Coherence Transform", Proceedings of the IEEE , vol.61, no.10, pp.1497,1498, Oct. 1973

[8] O. Hoshuyama, A. Sugiyama, A. Hirano: A Robust Adaptive Beamformer for Microphone Arrays with a Blocking Matrix using Constrained Adaptive Filters, IEEE Trans. Signal Process., vol. 47, no. 10, pp. 2677, 2684, 1999.